

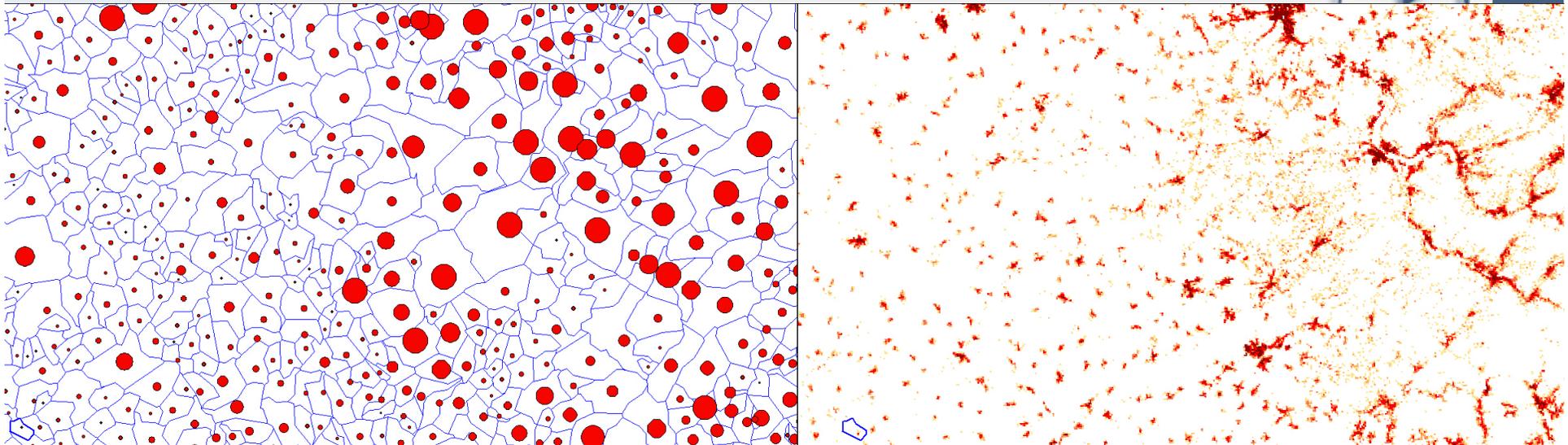


# Diffusion de données par carreau

## Les projets de l'INSEE

Jean-Luc LIPATZ  
INSEE





## La diffusion de données carroyées

**Préconisation 6** : mettre en place les conditions qui permettront aux utilisateurs de rassembler des **informations « clés »** issues de sources diverses selon une même maille géographique, le « carreau » ; ceci **afin de les agréger et de les analyser** sur leurs propres terrains infracommunaux d'observation et d'action.

# La diffusion de données carroyées

## Les questions



L'Insee et la statistique publique

Thèmes

Bases de données

Publications et services

Régions

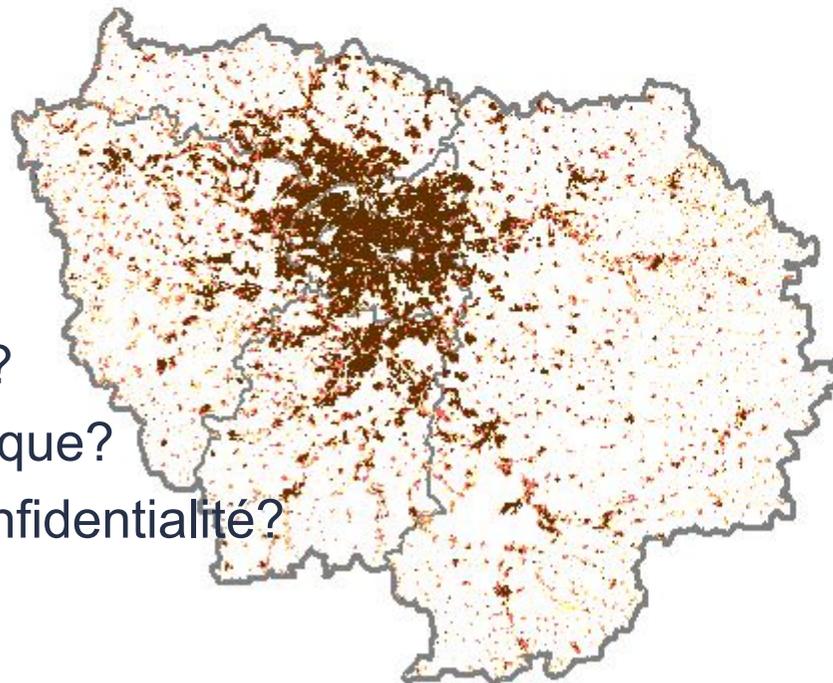
Définitions et méthodes

### Thèmes

- ▣ Agriculture
- ▣ Commerce
- ▣ Comptes nationaux - Finances publiques
- ▣ Conditions de vie - Société
- ▣ Conjoncture
- ▣ Économie
- ▣ Enseignement - Éducation
- ▣ Entreprises
- ▣ Industrie - IAA - Construction
- ▣ Population
- ▣ Revenus - Salaires
- ▣ Santé
- ▣ Services - Tourisme - Loisirs
- ▣ Territoire
  - ▣ Pays, villes d'Europe et du monde
  - ▣ Régions, départements et

▣ Accueil ▣ Thèmes ▣ Territoire ▣ Régions, dé... ▣ Données car... ▣ Données à 2... ▣ Ile de Fra

▣ Ile de France



### :: Légende

Densité de pop  
Nombre d'habitants  
par carreaux de

0 - 4
5 - 9
10 - 19
20 - 59
>= 60

### :: Pour en savoir

- ▣ Documentati  
données carr

### :: Liens

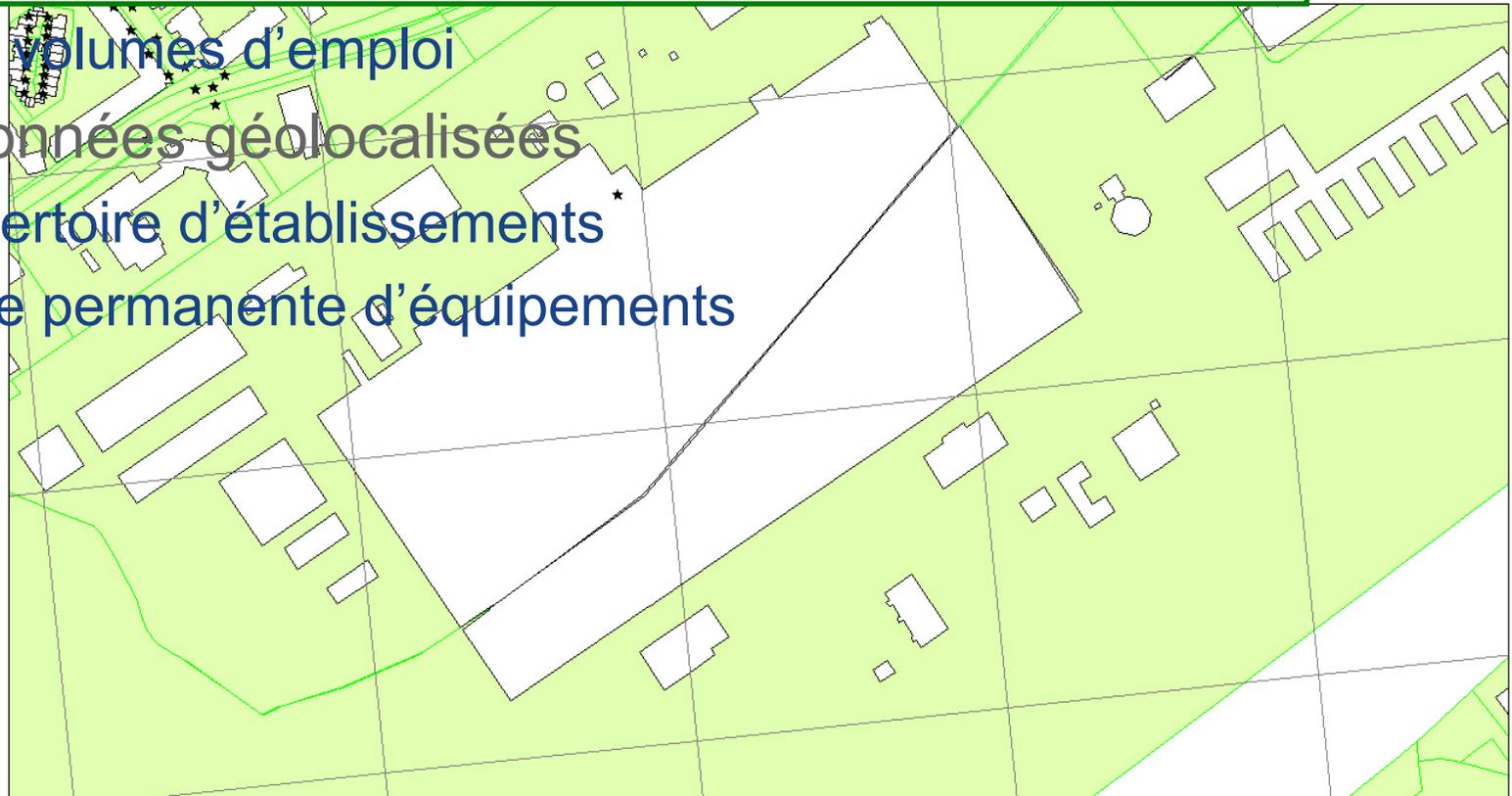
- ▣ Changer de r

- 1- Quel mode de diffusion?
- 2- Quelle maille géographique?
- 3- Comment assurer la confidentialité?
- 4- Quelles données?
- 5- Quel agenda?



## La cible : un dispositif double

- › Des données carroyées
  - Logements, ménages et population
  - Des volumes d'emploi
- › Des données géolocalisées
  - Répertoire d'établissements
  - Base permanente d'équipements





# 1- Le mode de diffusion

- › Diffusion grand public sur insee.fr
  - Indicateurs généraux
  - N'exclut pas la mobilisation d'autres indicateurs dans des contextes plus spécifiques (partenariats)
  
- › Insertion à trouver dans un cadre plus large
  - Ce sont des données spatialisées (INSPIRE?)
  - Les données de population ont déjà été copiées faute d'insertion dans un dispositif plus moderne (méta données, WMS?)



## 2- La géographie

- › Le champ : les carroyages sont destinés à transgresser toutes les frontières
  - toutes les communes de France
  - Métropole : quelle que soit la taille de la commune; il n’y aura pas de diffusion sur des champs partiels
    - Portées manquantes : Ile de Sein, Ile de Molène
  - DOMs
    - Réunion OK
    - Martinique partielle
    - Guadeloupe, Cayenne, Mayotte : rien pour l’instant
  
- › La maille géographique : il y aurait pu y avoir concertation, mais...
  - mailles de 200m de coté
  - ... l’Europe pose le LAEA comme standard de projection
  - ... l’INSEE est parti sur 200m de coté devenu standard de fait (AUs)

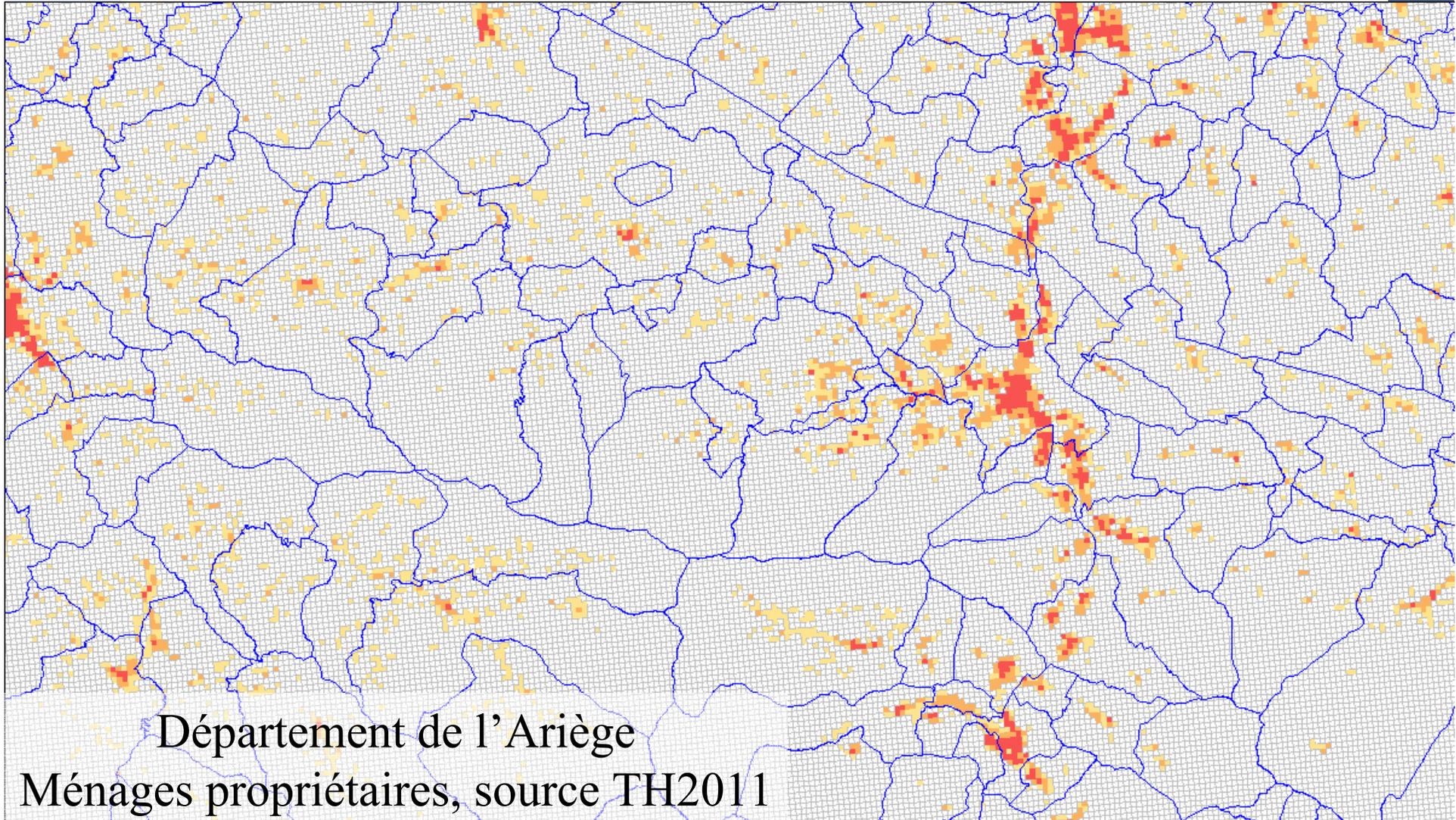
Page 6





## 2- La géographie

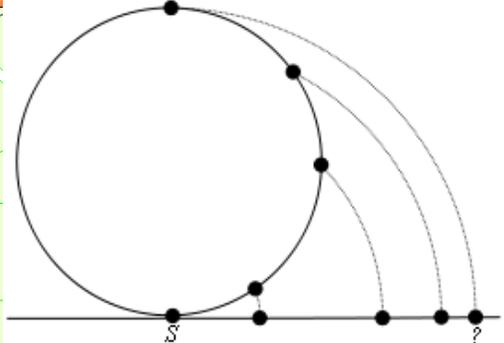
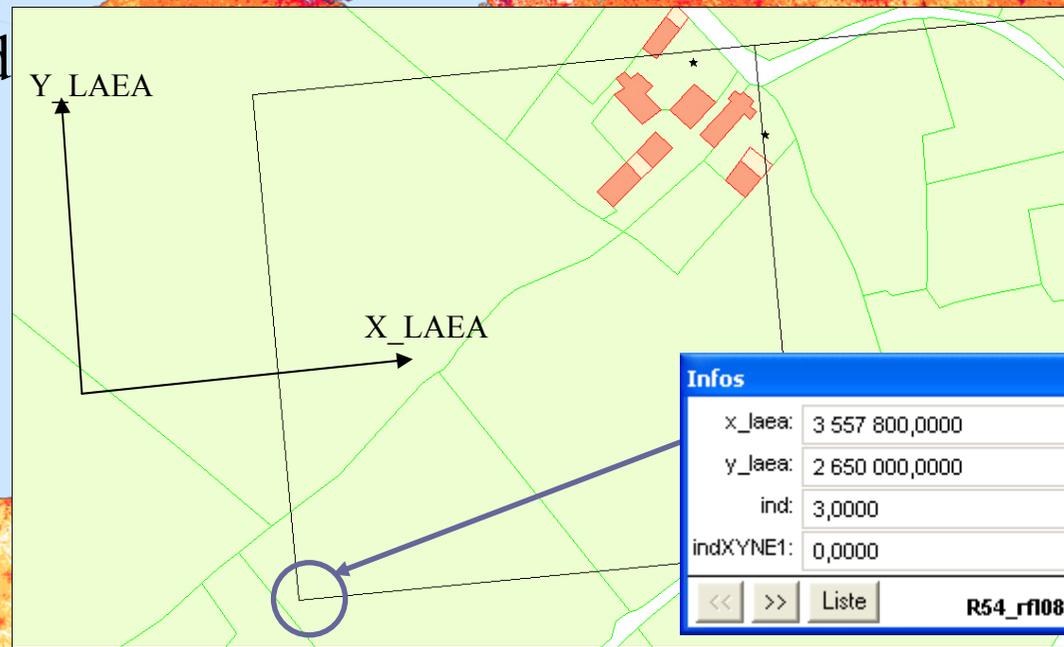
### Pas de limite : on a retrouvé Suzan!



## 2- La géographie Le Lambert Azimutal Equal Area (LAEA) Des carreaux pas carrés

S=Hanovre

Nord



Infos

x_jeaa:	3 557 800,0000
y_jeaa:	2 650 000,0000
ind:	3,0000
indXYNE1:	0,0000

<< >> Liste R54\_rf08\_LAEA200

Extrait de la base MIF/MID de diffusion, résolution 200m

Coordonnées LAEA du coin Sud-Ouest  
Population fiscale  
Population fiscale non localisée à la parcelle



- + Environnement
- Travail - Emploi

### :: Téléchargement des données

Chaque fichier compacté (format zip) contient les données sous deux formes :

- Un fichier en format dBase (DBF),
- Un fond de carte en format de transport MapInfo (MIF/MID).

Dans les deux cas, les carreaux décrits dans les fichiers sont uniquement ceux portant une population non nulle.

 [Télécharger les données](#) (format zip : 2 412 Ko)

### :: Téléchargement les fonds de carte départementaux complets

Chaque fichier zippé (\*.zip) contient le fonds de carte départemental complet en format de transport MapInfo (MIF/MID).

Les données de population sont proposées dans le fichier régional ci-dessus.

 [Télécharger la documentation](#)

 [Télécharger le fonds de carte département 75](#) (format zip : 64,5 Ko)

[le département 77](#) (format zip : 4,3 Mo)

[le département 78](#) (format zip : 1,5 Mo)

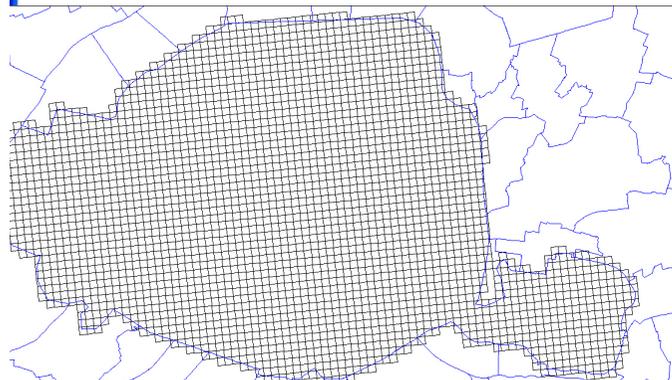
[le département 91](#) (format zip : 1,1 Mo)

[le département 92](#) (format zip : 112 Ko)

[le département 93](#) (format zip : 147,5 Ko)

[le département 94](#) (format zip : 151 Ko)

[le département 95](#) (format zip : 748 Ko)



**Infos**

x\_jea:

y\_jea:

COG1:

COG2:

COG3:

COG4:

COG5:

COG6:

id:

<< >> Liste dep\_75

# Pourquoi 200m?

Saint-Michel sur Orge 91570

RIL, cadastre et carroyage à 200m

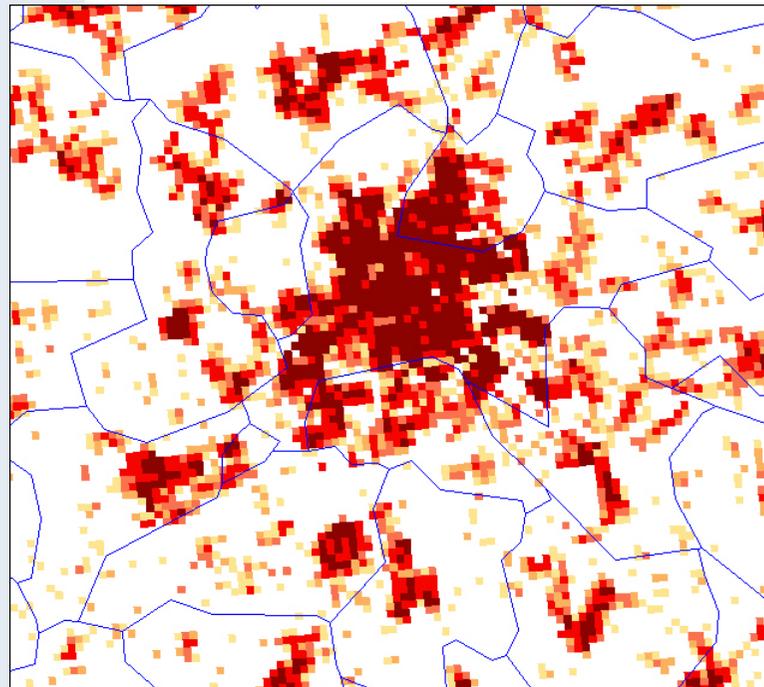


Un carroyage n'est pas un zonage  
C'est un positionnement approché



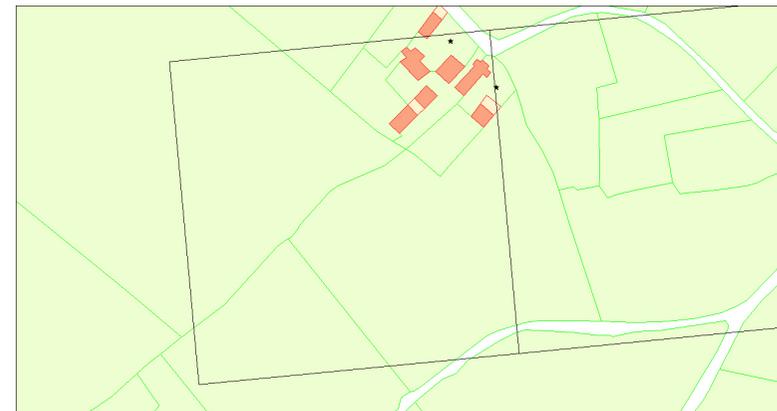
# Les carreaux sont les pixels d'une image

L'important c'est cela



Pas cela!

Livré sans garantie d'exactitude





### 3- Le traitement de la confidentialité

#### C'est l'image globale qui est importante

- › Objectif : garantir qu'aucune donnée individuelle ne sera accessible ni directement ni par approximation fiable.
- › La consultation de la CNIL se fera source par source
- › Définition de **populations à risque** (à faire par variable diffusée), exemple sur le revenu médian:
  - 1) Moins de 11 ménages
  - 2) Maille isolée ou ensemble de mailles isolées de moins de 80 ménages
  - 3) Ménages des mailles appartenant à un zonage standard (IRIS ou ZUS) de moins de 200 ménages
  - 4) Ménages des mailles ou le taux de bas revenus est extrême par rapport à la situation communale (« outliers »)
- › Traitement : **Permutation** des ménages avec des ménages de mailles voisines, si les caractéristiques sont différentes, **imputation** d'un revenu moyen sinon

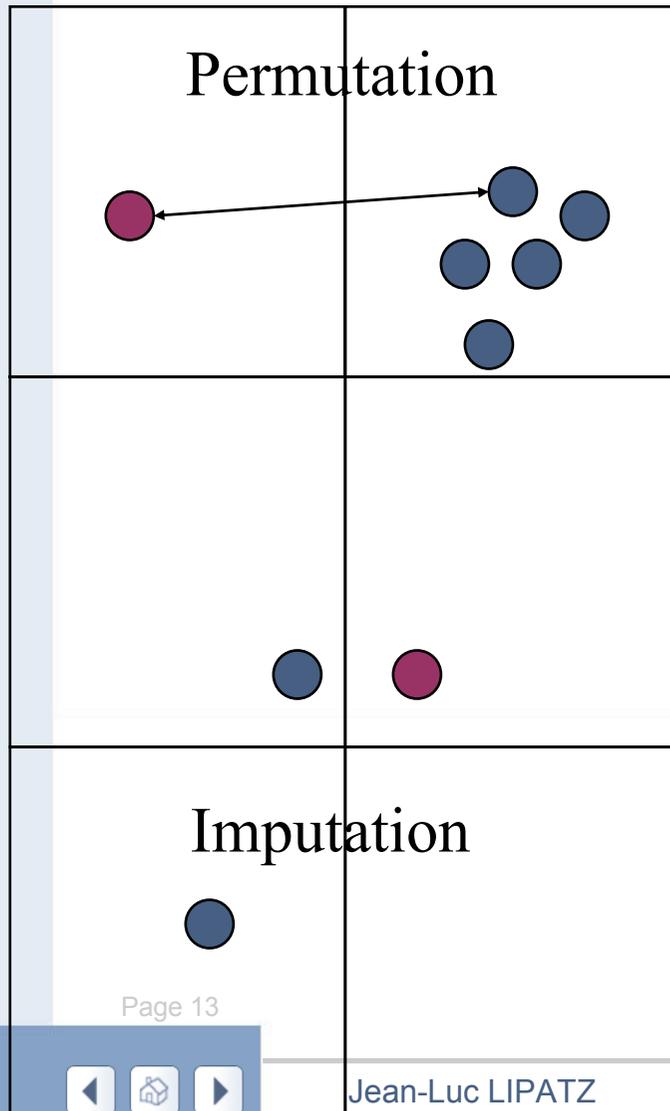




## Les données

Ne décriront pas la situation « exacte »  
carreau par carreau

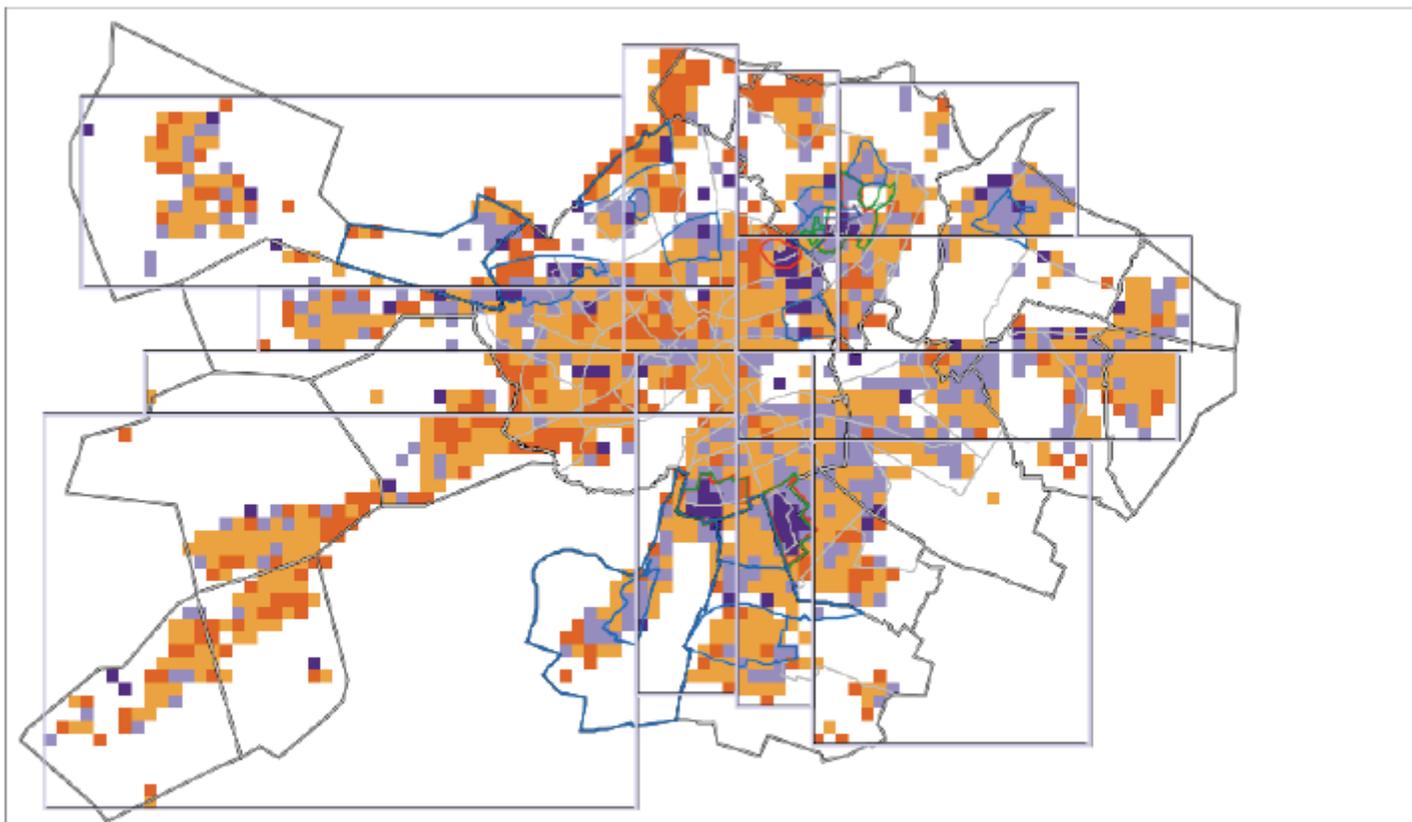
Mais reproduiront la situation exacte  
sur un sous ensemble suffisamment large





## Portée géographique du brouillage

- › Swapping : 9 plus proches carreaux
- › Imputation : Zones ad hoc pour disposer d'une population suffisante pour faire l'imputation



- › Mais toutes les règles ne seront pas publiées!



# Des données [garanties] fausses...

## Données originales :

id	n	SEXE_1	SEXE_2	AGE_1	AGE_2	AGE_3
00134	1	1	1	0	0	1
00135	1	1	1	0	0	1
00222	20	8	12	3	13	4
00308	11	8	3	3	8	0
00309	12	7	5	0	11	1
00394	3	2	1	0	1	2
00395	3	1	2	1	0	2
00473	1	0	1	0	1	0
00474	3	1	2	0	3	0

Brouillées

Pas de changement : hors sujet confidentialité

## Données publiables :

id	n	SEXE_1	SEXE_2	AGE_1	AGE_2	AGE_3
00134	1	0	1	0	0	1
00135	1	0	1	0	1	0
00222	20	11	9	4	12	4
00308	11	6	5	2	9	0
00309	12	6	6	0	11	1
00394	3	3	0	0	1	2
00395	3	1	2	0	1	2
00473	1	0	1	0	1	0
00474	3	2	1	0	3	0

Pas de changement (mais qui le saura?)

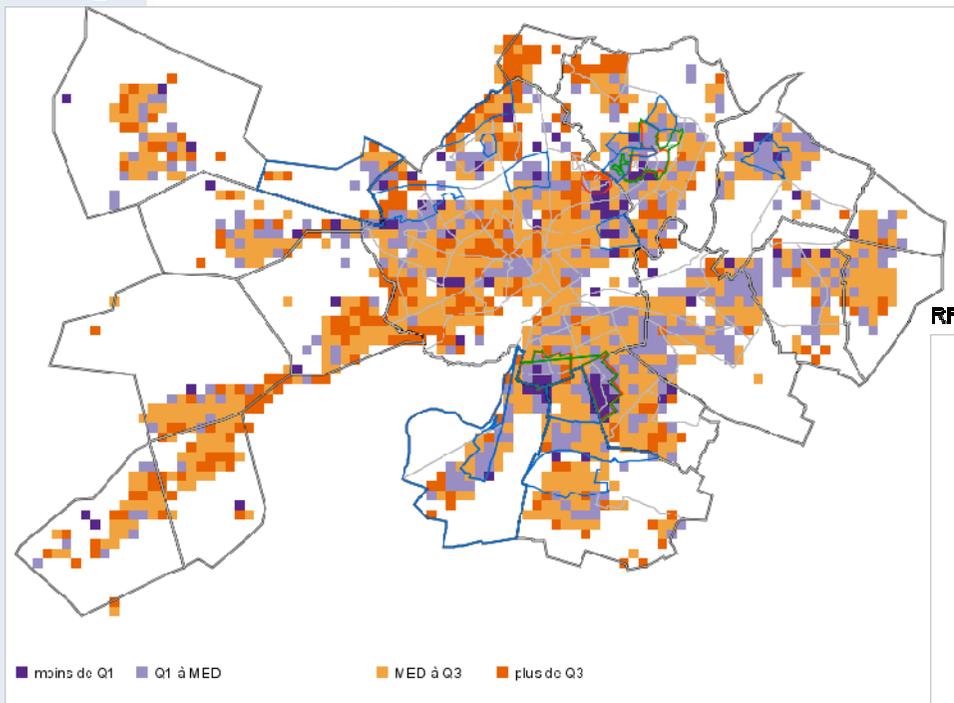




# ...mais dont l'utilisation reste possible

## Revenu médian par carreau, agglomération de Caen

RFL2009\_Caen - RFL - Revenu médian par carreau



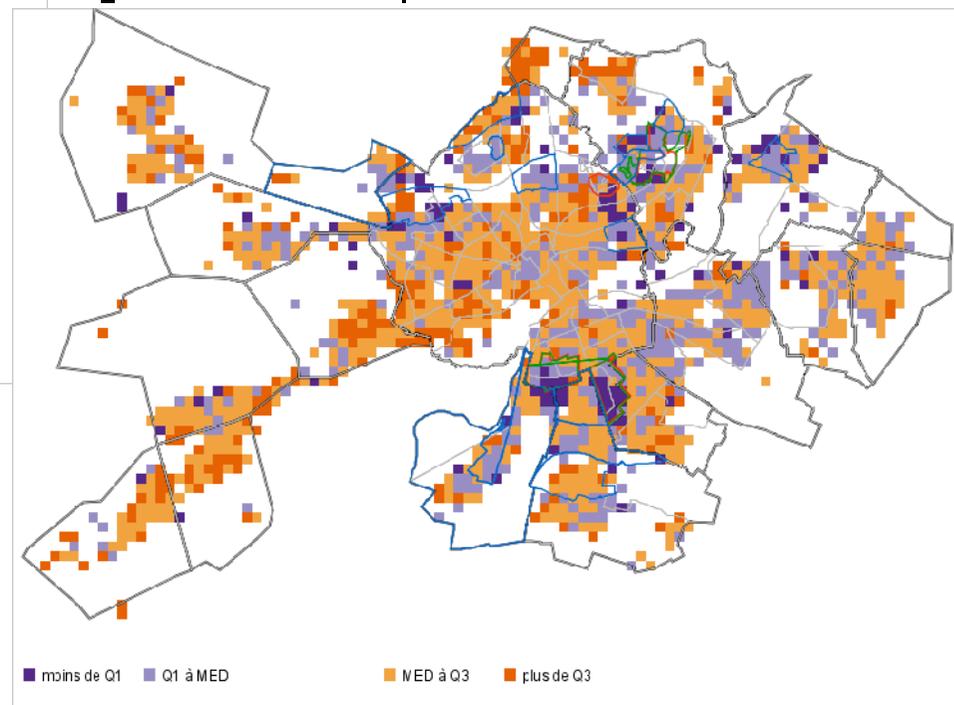
Données originales

Entre les deux :  
10 000 ménages/80 000 brouillés

Page 16

Données brouillées

RFL2009\_Caen - RFL - Revenu médian par carreau - Données traitées





## 4 – Les indicateurs

### Les trois filières de la géolocalisation

- › Reconnaissance de libellés d'adresse
  - Sources diverses : CNAM, PE, DADS, CAF...
  - Sous réserve de traitements manuels, plus ou moins lourds et difficiles selon les sources
  - ➔ A moyen terme – en fonction de l'intensité de la demande
- › Appariement au cadastre
  - Sources fiscales : TH, impôt sur le revenu, MAJIC
  - Complètement automatique
  - ➔ Immédiat
- › Recensement : estimations et appariement au bâtiment habité
  - Pas de problème sur les moins de 10 000 hors variables du complémentaire (questions de précision à outiller)
  - Production lourde sur les plus de 10 000, questions de précision à outiller, estimation variable par variable
  - ➔ A court terme

## 4- Les indicateurs (1+30)

# Ménages

Logement	2	Résidences principales, résidences secondaires
Population	9	<b>Totale</b> , par âge : [0,3], [0-5], [0,10], [0,14], [0,17], 25 et +, 65 et +, 75 et +
Ménages	2	D'une seule personne, de cinq personnes et plus
	2	Ménages propriétaires, locataires
	2	Ménages en individuel, en collectif
	1	Ménages présents depuis au moins cinq ans
	1	Surface cumulée des logements occupés
Revenus	2	Revenu par UC médian et nombre d'UCs
	2	Ménages de revenu/UC en dessous d'un seuil, au dessus d'un seuil
Activité	1	<b>Ménages bi-actifs</b>
	1	<b>Population active totale</b>
	3	<b>Employés, ouvriers, cadres</b>
	1	DEFM
	2	<b>Actifs occupés dont le mode de transport est la voiture, les TC</b>



## 5- Ebauche de calendrier (sous réserves administratives diverses)

TH+IRPP et RFL  
RP (estimations mixtes  
+pseudo RIL)  
Pôle emploi (Geoloc)

- › Fin 2012 : à partir des données disponibles
  - données fiscales
- › Début 2013 : après estimations
  - données issues du recensement
- › Mi 2013 : après géolocalisation
  - données de pôle emploi
- › Des compléments (à partir d'autres sources)?
- › Actualisation cinq ans après?

Page 19





**Merci de votre attention**

